It’s the Difference That Matters: An Argument for Contextually-Grounded Acoustic Intonational Phonology

Sasha Calhoun
Centre for Speech Technology Research
University of Edinburgh, UK
Sasha.Calhoun@ed.ac.uk

Linguistics Society of America Annual Meeting
Oakland, CA, 6 January 2005
Overview

- what makes a good intonational phonology?
- issues with ToBI: description, explanation, verification
- pilot study: using linear regression modelling to find acoustic cues to topic status
- intensity, duration, phrase level and relative $f_0$ cues all significant
- suggest phonological investigations should be corpus-based, categories are bundles of weighted acoustic features affected by context
Qualities of a good phonology

= the structure of supra-segmental speech sounds

- describe parts of speech signal relevant to the conveyance of intonation categories
- explain how these intonation categories convey meaning
- be verifiable
- give coverage of differences between languages and varieties of one language
Intonational Phonology & ToBI

(Silverman et al 92, Pierrehumbert & Hirschberg 90)

- discourse meanings = status of entities in speaker and hearer’s mutual discourse model
- perceptual intonation categories: pitch accents and boundary tones
- prosodic output = classes of pitch contours
- ‘paralinguistic’ connotations, effect of hierarchical structure

<table>
<thead>
<tr>
<th>F0</th>
<th>ToBI</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>H*</td>
<td>L+H*</td>
<td>L</td>
</tr>
<tr>
<td>L*</td>
<td></td>
<td>H%</td>
</tr>
</tbody>
</table>

= meaning

THAT moves the SQUARE
Issues

- **description:**
  - bias on **local features** of pitch contour (c.f. importance of relative height: Gussenhoven & Rietveld 88, Terken 91, Ladd 96; \( f0 \) on unstressed syllables: Xu et al 04)
  - bias on **\( f0 \) turning points** (c.f. other factors which affect alignment: van Santen & Möbius 00, segmental effects; Scherer et al 04, emotions)
  - no explicit modelling of effect of other **layers of structure**

- **linear regression modelling**
- **bundles of acoustic features**
- **include previous context as features**
Issues

**explanation:**
- very **uneven distribution** of pitch accents (Taylor 00: 80% H*)
- little evidence of emerging consensus on **pitch accent meanings** (e.g. status of L+H*, H* P&H 90, Steedman 00, Lambrecht & Michaelis 98 all differ, Hedberg & Sosa 01 corpus - mixed)

**verification:**
- **low inter-annotator agreement** on pitch accent types (Silverman et al 92: 61%)
- difficult to find criteria to confirm or reject existence of **perceptual categories**

**start with meanings**

**corpus-based research**
Topic Status Experiment with SPOT Corpus

- investigate intonational marking of topics in SPOT corpus
- tested given, new and contrastive categories
- SPOT = dialogues collected as part of a game task by Schafer (Hawaii), Speer (Ohio), Warren (Victoria, NZ) and colleagues
- used 52 utterances involving 16 pairs of male speakers of American English
- these utterances ToBI transcribed as part of the original project
Discourse Contexts

- tested the realisation of the word *square* in different discourse contexts in a game task where people had to, among other things, move squares with cylinders.

Q: Which cylinder do you want to change the position of the square?
A: The red one. When that moves the *square*, it should land in a good spot.

*given topic*

Q: Which cylinder do you want to change the position of this time?
A: The red one. When that moves the *square*, it should land in a good spot.

*new topic*

Q: (I know which cylinder you want to change the position of the triangle, ) but which cylinder do you want to change the position of the square?
A: The red one. When that moves the *SQUARE*, it should land in a good spot.

*contrastive topic*
ToBI - No clear mapping

- no statistically significant relationship between ToBI pitch accents and topic status

<table>
<thead>
<tr>
<th></th>
<th>Ø</th>
<th>L*</th>
<th>H*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Given</strong></td>
<td>4 (44%)</td>
<td>1 (11%)</td>
<td>4 (44%)</td>
</tr>
<tr>
<td><strong>New</strong></td>
<td>6 (19%)</td>
<td>6 (19%)</td>
<td>19 (61%)</td>
</tr>
<tr>
<td><strong>Contrastive</strong></td>
<td>4 (33%)</td>
<td>1 (18%)</td>
<td>7 (58%)</td>
</tr>
</tbody>
</table>
‘Bundles’ of Acoustic Cues

- using a linear regression model, the $f_0$ mean of the and square, and the duration of square were all significant predictors of topic status:

<table>
<thead>
<tr>
<th>words</th>
<th>F0 (semitones base 100 Hz)</th>
<th>$\text{square } f_0$ mean (semitones base 100 Hz)</th>
<th>$\text{square }$ duration (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>given</td>
<td>-0.72</td>
<td>-1.11</td>
<td>365</td>
</tr>
<tr>
<td>new</td>
<td>1.71</td>
<td>1.23</td>
<td>423</td>
</tr>
<tr>
<td>contrastive</td>
<td>0.922</td>
<td>2.08</td>
<td>463</td>
</tr>
</tbody>
</table>
however, we find that topic status can also be predicted using a linear regression model with acoustic features of the preceding utterance as features.

<table>
<thead>
<tr>
<th></th>
<th>that f0 mean</th>
<th>moves f0 mean</th>
<th>moves mean intensity (dB)</th>
<th>phrase f0 mean</th>
<th>phrase mean intensity (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>given</td>
<td>0.0</td>
<td>1.4</td>
<td>67</td>
<td>-0.6</td>
<td>75</td>
</tr>
<tr>
<td>new</td>
<td>2.6</td>
<td>3.2</td>
<td>65</td>
<td>1.6</td>
<td>72</td>
</tr>
<tr>
<td>contrastive</td>
<td>2.5</td>
<td>3.6</td>
<td>65</td>
<td>2.1</td>
<td>74</td>
</tr>
</tbody>
</table>
Relative Acoustic Cues

- topic status can also be signalled by the level of different acoustic cues on square relative to their value in the preceding utterance.

<table>
<thead>
<tr>
<th>Words</th>
<th>that - square f0 difference</th>
<th>moves - square f0 difference</th>
<th>moves - square intensity difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>given</td>
<td>1.1</td>
<td>2.5</td>
<td>- 0.5</td>
</tr>
<tr>
<td>new</td>
<td>1.3</td>
<td>2.0</td>
<td>- 0.6</td>
</tr>
<tr>
<td>contrastive</td>
<td>0.4</td>
<td>1.5</td>
<td>- 1.9</td>
</tr>
</tbody>
</table>
Summary - Topic Marking

- intensity and duration are significant cues to intonation categories along with $f_0$
- ‘given’ versus ‘new’/‘contrastive’ topic status appear to be features of whole intonation phrases
- ‘new’ versus ‘contrastive’ topic status could be marked by the relative $f_0$ height and intensity of square compared to that and moves
Conclusions and Research Directions

- seeing intonational categories as bundles of weighted acoustic features allows statistical modelling of intonational phonology
- this increases **descriptive** power and makes the model **verifiable**
- the approach also **explains** how meaning is conveyed much more transparently
- statistical variation explicitly models contextual variation
- larger studies on these lines allow a more principled way to discover the apparatus of intonation, e.g. pitch accents, branching structure, phrase properties, etc.